

IN THE TITLE:

Please delete the title of the invention and replace it with the following title:

Methods for Extracting Information on Interactions Between Biological Entities From  
Natural Language Text Data.

IN THE SPECIFICATION:

On page 1, delete the first paragraph and replace it with the following paragraph:

This application is a continuation-in-part of pending application Serial No.

09/327,938 filed June 8, 1999 which claims priority to provisional patent application  
Serial No. 60/129,469 filed April 15, 1999. The invention described herein was funded  
in part by a grant from the National Library of Medicine, namely, Grant Number's  
LM06274 and LM05627. The United States Government may have certain rights to the  
invention. The present specification contains a computer program listing which appears  
as a microfiche Appendix H.

On page 34-35, delete the last paragraph which continues on page 35 and replace  
it with the following paragraph:

Known motifs/domains for proteins may also be collected using the flat file  
versions of major protein databases, such as SwissProt and the non-redundant database  
of NCBI. The databases can be downloaded and searched for the keywords "motif" and  
"domain" in the feature tables of proteins. In addition, existing databases of motifs and  
domains, such as BLOCKS and pfam, can be downloaded (Henikoff et al., 1991, NAR  
19:6565-6572). Still further, it is understood that any publically available database

D2  
containing gene/protein sequences may be utilized to generate the specialized databases for use in the practice of the present invention.

---

On page 44, delete the first full paragraph and replace it with the following paragraph:

---

D3  
To construct a reconciled tree according to the invention, the first step comprises a search for homologs in a publicly or privately available database such as, for example, GenBank, Incyte, binary BLAST databases, Swiss Prot and NCBI databases. Following the identification of homologous sequences a global alignment is performed using, for example, the CLUSTALW program. From the sequence alignment a gene tree is constructed using, for example, the computer program CLUSTLAW which utilizes the neighbor-joining method of Saito and Nei (1997, Mol. Biol. Evol. 4:406-425).

Construction of a species tree is then retrieved from, for example, the following database  
3.NCBI.NLM.NIH.GOV//taxomy.tax.html.

---

On page 66, delete the first and second paragraph and replace them with the following paragraphs:

---

D4  
Identification of a putative apoptosis-related human gene began with an identification of all genes in *C. elegans* that contained either a POZ or kelch domain. A subset of these genes is shown in Figure 13. Hidden Markov Models (HMM) for the POZ and Kelch domains were built as follows. Starting with POZ and kelch sequences from the *Drosophila* kelch protein (gi 577275) homologs were identified in other protein sequences using the BLASTP program. The resulting sequences showing significant similarity (e-value less than 0.001) were aligned using CLUSTALW program and the alignments were used to build Hidden Markov Models with HMMER-2 package